EKSTRAKSI INFORMASI NAMA-NAMA MAKANAN NUSANTARA BERBASIS N-GRAM DAN LEXICON MODEL

Skripsi

oleh

EDO PRASETIA WIJAYA KHAIRIL

71130056

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS KRISTEN DUTA WACANA

2018

EKSTRAKSI INFORMASI NAMA-NAMA MAKANAN NUSANTARA BERBASIS N-GRAM DAN LEXICON MODEL

Skripsi

Diajukan kepada Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana
Sebagai Salah Satu Syarat dalam Memperoleh Gelar
Sarjana Komputer

Disusun oleh

EDO PRASETIA WIJAYA KHAIRIL
71130056

PROGRAM STUDI INFORMATIKA FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
2018

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

EKSTRAKSI INFORMASI NAMA-NAMA MAKANAN NUSANTARA BERBASIS N-GRAM DAN LEXICON MODEL

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 1 Agustus 2018

EDO PRASETIA WIJAYA KHAIRIL
71130056

HALAMAN PERSETUJUAN

Judul Skripsi	:	EKSTRAKSI INFORMASI NAMA-NAMA MAKANAN NUSANTARA BERBASIS N-GRAM DAN LEXICON MODEL
Nama Mahasiswa	:	EDO PRASETIA WIJAYA KHAIRIL
N I M	:	71130056
Matakuliah	:	Skripsi (Tugas Akhir)
Kode	:	TIW276
Semester	:	Genap
Tahun Akademik	:	2017/2018

Telah diperiksa dan disetujui di
Yogyakarta,
Pada tanggal 1 Agustus 2018

Dosen Pembimbing I	Dosen Pembimbing II
Lucia Dwi Krisnawati, Dr. Phil.	Gloria Virginia, S.Kom., MAI, Ph.D.

HALAMAN PENGESAHAN

EKSTRAKSI INFORMASI NAMA-NAMA MAKANAN NUSANTARA BERBASIS N-GRAM DAN LEXICON MODEL

Oleh: EDO PRASETIA WIJAYA KHAIRIL / 71130056

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 26 Juli 2018

Yogyakarta, 1 Agustus 2018
Mengesahkan,

Dewan Penguji:

1. Lucia Dwi Krisnawati, Dr. Phil.
2. Gloria Virginia, S.Kom., MAI, Ph.D.
3. Danny Sebastian, S.Kom., M.M., M.T.
4. Hendro Setiadi, M.Eng

Dekan	Ketua Program Studi
(Budi Susanto, S.Kom., M.T.)	(Gloria Virginia, Ph.D.)

UCAPAN TERIMA KASIH

Dalam penelitian tugas akhir ini, penulis mendapatkan bantuan, saran, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Budi Susanto, S.Kom., M.T. selaku Dekan Fakultas Teknologi Informasi Universitas Kristen Duta Wacana.
2. Ibu Gloria Virginia, S.Kom., MAI., Ph.D. selaku Ketua Program Studi Teknik Informatika Universitas Kristen Duta Wacana.
3. Ibu Dr. Lucia Dwi Krisnawati, selaku dosen pembimbing I yang telah memberikan waktu secara rutin untuk melakukan konsultasi dan memberikan saran dan masukan mengenai pemrograman sistem, dan penyelesaian masalah dengan cara yang lebih sederhana.
4. Ibu Gloria Virginia, S.Kom., MAI., Ph.D. selaku dosen pembimbing II yang telah meluangkan waktu untuk memberikan bimbingan dan memberikan saran serta masukan mengenai penulisan laporan juga analisis sistem.
5. Orangtua, saudara-saudara dan teman-teman terdekat yang selalu memberikan dukungan motivasi kepada penulis untuk menyelesaikan tugas akhir ini.
6. Pihak – pihak lain yang tidak dapat penulis sebutkan satu per satu yang berperan secara langsung maupun tidak langsung selama pengerjaan tugas akhir.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan kasih karunia-Nya sehingga penulis dapat menyelesaikan pembuatan sistem dan laporan tugas akhir dengan judul "EKSTRAKSI INFORMASI NAMA-NAMA MAKANAN NUSANTARA BERBASIS *N-GRAM* DAN *LEXICON MODEL*" dengan baik.

Penulisan laporan tugas akhir ini diajukan sebagai salah satu syarat guna mencapai gelar Sarjana Strata Satu (S1) di Fakultas Teknologi Informasi Program Studi Teknik Informatika Universitas Kristen Duta Wacana Yogyakarta.

Dalam pembuatan laporan ini, penulis menyadari bahwa masih ada kekurangan, baik dari materi maupun teknik penyajiannya. Oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran dari pembaca. Akhir kata penulis memohon maaf apabila dalam penulisan laporan ini, ada kalimat yang kurang berkenan. Semoga hasil dari pengerjaan tugas akhir ini dapat berguna dan bermanfaat bagi banyak pihak.

Yogyakarta, 1 Agustus 2018

Penulis

INTISARI

EKSTRAKSI INFORMASI NAMA-NAMA MAKANAN NUSANTARA BERBASIS *N-GRAM* DAN *LEXICON MODEL*

Ekstraksi Informasi merupakan proses mengubah teks tidak terstruktur menjadi informasi dalam bentuk terstruktur. Tujuan dari perancangan sistem ini adalah untuk membantu proses Ekstraksi Informasi nama-nama makanan terhadap dokumen tentang makanan nusantara. Ekstraksi Informasi nama-nama makanan dapat dilakukan dengan berbagai metode, yaitu metode berbasis *n-gram model*, metode *lexicon model*, metode berbasis aturan, metode *syntactic analysis*, metode *pattern discovery* dan metode *name entity recognition* dan metode *relation extraction*.

Pada penelitian ini, penulis menggunakan metode berbasis *n-gram*, leksikon dan dilengkapi dengan metode berbasis aturan untuk mengekstraksi nama-nama makanan nusantara. *Preprocessing* dalam penelitian ini menggunakan normalisasi, tokenisasi, dan *case folding*. Proses pengecekan dilakukan dengan menggunakan aturan-aturan, seperti aturan penetapan nilai *threshold*, aturan seleksi *bigram*, dan aturan penggabungan *bigram*. Hasil *output* sistem berupa daftar nama makanan dan proses perhitungan sistem dalam mengekstraksi nama makanan.

Hasil menunjukan rata-rata nilai akurasi sebesar 97% yang mengindikasikan metode berbasis *n-gram*, *lexicon* model, dan aturan cukup baik dalam mengekstraksi nama makanan. Kelemahan sistem adalah sistem belum bisa mengekstraksi kata dalam nama makanan yang bermakna bukan makanan, sehingga hasil yang terbentuk tidak sesuai dengan konteks makna dalam kalimat.

Kata Kunci: [Ekstraksi Informasi, *N-gram*, leksikon, berbasis aturan]

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

INTISARI

EKSTRAKSI INFORMASI NAMA-NAMA MAKANAN NUSANTARA BERBASIS *N-GRAM* DAN *LEXICON MODEL*

Ekstraksi Informasi merupakan proses mengubah teks tidak terstruktur menjadi informasi dalam bentuk terstruktur. Tujuan dari perancangan sistem ini adalah untuk membantu proses Ekstraksi Informasi nama-nama makanan terhadap dokumen tentang makanan nusantara. Ekstraksi Informasi nama-nama makanan dapat dilakukan dengan berbagai metode, yaitu metode berbasis *n-gram model*, metode *lexicon model*, metode berbasis aturan, metode *syntactic analysis*, metode *pattern discovery* dan metode *name entity recognition* dan metode *relation extraction*.

Pada penelitian ini, penulis menggunakan metode berbasis *n-gram*, leksikon dan dilengkapi dengan metode berbasis aturan untuk mengekstraksi nama-nama makanan nusantara. *Preprocessing* dalam penelitian ini menggunakan normalisasi, tokenisasi, dan *case folding*. Proses pengecekan dilakukan dengan menggunakan aturan-aturan, seperti aturan penetapan nilai *threshold*, aturan seleksi *bigram*, dan aturan penggabungan *bigram*. Hasil *output* sistem berupa daftar nama makanan dan proses perhitungan sistem dalam mengekstraksi nama makanan.

Hasil menunjukan rata-rata nilai akurasi sebesar 97% yang mengindikasikan metode berbasis *n-gram*, *lexicon* model, dan aturan cukup baik dalam mengekstraksi nama makanan. Kelemahan sistem adalah sistem belum bisa mengekstraksi kata dalam nama makanan yang bermakna bukan makanan, sehingga hasil yang terbentuk tidak sesuai dengan konteks makna dalam kalimat.

Kata Kunci: [Ekstraksi Informasi, *N-gram*, leksikon, berbasis aturan]

BAB 1
PENDAHULUAN

1.1. Latar Belakang

Masakan indonesia (nusantara) merupakan salah satu tradisi kuliner yang paling kaya di dunia. Kekayaan ini merupakan cermin dari keberagaman budaya dan tradisi dari sekitar 6.000 pulau berpenghuni (Day, 2015). Masakan nusantara kaya dengan bumbu yang berasal dari rempah-rempah dan penuh dengan cita rasa yang kuat. Oleh karena itu, masakan nusantara memiliki aneka ragam nama masakan baik yang dipengaruhi dari berbagai daerah di nusantara atau dipengaruhi dari luar negeri.

Kemajuan teknologi dalam berkomunikasi mengakibatkan informasi seputar masakan nusantara dapat ditemukan oleh siapa saja. Dalam beberapa tahun terakhir berbagai macam teknologi menciptakan cara baru berkomunikasi dalam berbisnis. Cara baru dalam berkomunikasi ini disebut *Social Communication Model* (Hanna, 2009). Model tersebut menciptakan cara berkomukasi yang interaktif dan terbuka bagi siapa saja yang ingin berpartisipasi. *Social Communication Model* menciptakan trend "let's have a conversation" dan menghilangkan trend lama "we talk, you listen" (Bovée, 2016). Dengan kemajuan teknologi ini, informasi seputar nama masakan nusantara beredar luas dan cepat.

Para pelaku bisnis masakan memanfaatkan model tersebut untuk memantau produk-produk masakan yang dijual. Dampak diterapkannya model ini adalah informasi seputar masakan beredar bebas, tidak terbatas dan bahkan dapat dihasilkan oleh siapa saja yang menggunakan berbagai media yang mendukung model komunikasi tersebut. Hal ini menyebabkan "Information Overload" dimana para pelaku bisnis menerima informasi lebih banyak dari yang dapat diproses lebih lanjut (Craig, 2008) dan dapat menyebabkan "Information Technology Paradox" dimana media komunikasi yang dipakai yaitu teknologi menjadi tidak berfungsi sebagaimana mestinya (24/7 Wall St, 2010).

Ekstraksi informasi merupakan suatu bidang ilmu dalam pengolahan bahasa alami, yang mengubah *teks* tidak terstruktur menjadi informasi dalam bentuk terstruktur (Susanti, 2015). *Teks* ini ditransmisikan secara tidak terstruktur melalui *Internet* seperti *website*, termasuk teks yang membahas seputar masakan nusantara. Oleh karena itu, Mengekstraksi informasi menjadi hal penting (Scheffer, 2002) untuk dilakukan dalam mengidentifikasi nama-nama masakan nusantara dengan informasi lainnya.

Banyak metode ekstraksi informasi yang dapat digunakan untuk mengekstrasi misalnya metode *N-Gram* dan *Lexicon Model*. *N-gram Model* adalah model probabilistik yang awalnya dirancang oleh ahli matematika dari Rusia pada awal abad ke-20 dan kemudian dikembangkan untuk memprediksi item berikutnya dalam urutan item. Item tersebut bisa berupa huruf / karakter, kata, atau yang lain sesuai dengan aplikasi. Dalam memprediksi peneliti membutuhkan bantuan corpus (plural corpora). Corpus adalah kumpulan koleksi dari *text* yang dapat dibaca computer. *Lexicon Model* (corpus) yang digunakan dalam penelitian ini berupa text nama makanan nusantara.

1.2. Perumusan Masalah

Berdasarkan uraian di atas maka masalah yang akan diteliti adalah sebagai berikut:

1. Bagaimana sistem dapat mengidentifikasi nama-nama masakan nusantara?
2. Bagaimana sistem dapat mengekstraksi nama-nama masakan nusantara sehingga lebih mudah terbaca atau ditemukan?
3. Bagaimana akurasi dari sistem yang akan dibuat dalam melakukan ekstraksi nama-nama masakan nusantara?

1.3. Batasan Masalah

Berikut batasan masalah dalam penelitian ini:

1. Sumber dokumen berasal dari media daring seperti *website*, *blog*, dan media sosial yang membahas makanan Indonesia.

2. Sumber nama makanan berasal dari media daring seperti *website*, *blog*, media sosial, dan menu makanan restoran Indonesia.
3. Dokumen latih dan leksikon hanya membahas tentang makanan Indonesia.
4. Keluaran akhir dari sistem berupa nama makanan nusantara.

1.4. Tujuan Penelitian

1. Mengembangkan perangkat lunak yang mampu mengekstraksi nama-nama makanan.
2. Sistem ekstraksi informasi berbasis *web*.

1.5. Manfaat Penelitian

Berikut manfaat penelitian dalam tugas akhir ini:

1. Memudahkan pengidentifikasian nama – nama masakan nusantara yang terdapat pada artikel pada *website*, *blog*, dan media sosial seperti *facebook*.
2. Mengetahui tingkat kecocokan metode *N-Gram Model* dan *Lexicon Model* dalam mengidentifikasikan nama – nama masakan nusantara.
3. Untuk penelitian lebih lanjut sistem ini dapat digunakan bersama dengan sistem yang lebih kompleks seperti Sentimen Analisis untuk mengetahui jenis komentar (positif atau negatif) terhadap makanan nusantara.

1.6. Metode Penelitian

1.6.1. Tahap Pengumpulan dan Persiapan Data

Pengumpulan data dokumen latih dan dokumen uji yang digunakan diambil dari artikel *online* pada *website* yang berhubungan dengan masakan nusantara serta komentar yang mengandung nama makanan nusantara yang telah dipilih peneliti dan disimpan dalam bentuk *database*. Dokumen latih sebanyak 1000 buah, dokumen uji sebanyak 30 buah dan dalam leksikon terdapat 1000 nama masakan nusantara.

Tahap selanjutnya adalah tahap persiapan data, dokumen tersebut akan dipecah menjadi kata-kata. Pemecahan dokumen menjadi kata-kata tersebut akan melalui pra-pemrosesan terlebih dahulu. Hasil dari pra-pemrosesan ini adalah kata yang akan dijadikan acuan dalam melakukan identifikasi nama masakan nusantara pada data uji yang akan diteliti. Kata tersebut nantinya akan dihitung frekuensi kemunculannya dalam suatu dokumen latih.

1.6.2. Tahap Implementasi Sistem

Pada tahap implementasi ini, penelitian akan diterapkan pada sebuah *web*. Pengguna akan memasukkan dokumen yang akan diuji, dan *web* akan menganalisa setiap kata dari dokumen tersebut untuk mencari nama makanan nusantara dari suatu dokumen. *web* akan mencari nilai probabilitas *bigram* (jenis *N-Gram Model*) yang ditentukan dari nilai frekuensi kemunculan kata pada dokumen *training*. Setelah mendapatkan nilai probabilitas tertinggi dari perhitungan sebelumnya, maka dokumen tersebut dikategorikan berdasarkan kategori dengan probabilitas tertinggi. Probabilitas tinggi berarti mengindikasikan bahwa kata tersebut merupakan nama masakan nusantara. Hasil akhirnya *web* akan menebalkan kata yang menjadi nama masakan nusantara.

1.6.3. Tahap Pengujian Sistem

Dalam melakukan pengujian sistem kategorisasi, data yang akan digunakan adalah data dokumen training, dokumen uji dan korpus. Sedangkan *sample* yang digunakan untuk menguji sistem ini berjumlah 30 *sample*, dari 30 *sample* tersebut akan diproses oleh sistem dan dikategorikan oleh peneliti dokumen yang di identifikasi benar atau salah. Lalu pengujian dilanjutkan ke *confusion matrix*.

1.7. Sistem Penulisan

Bab 1 berisi pendahuluan terdiri dari latar belakang masalah, rumusan masalah,batasan masalah, tujuan penelitian, metode penelitian, dan sistem

penulisan dari judul yang telah diangkat yaitu “Ekstraksi Informasi Nama-Nama Makanan Nusantara berbasis *N-Gram* dan *Lexicon Model*.”

Bab 2, Tinjauan pustaka dan landasan teori, tinjauan pustaka akan membahas mengenai jurnal/*paper* yang berkaitan dengan penelitian tersebut, dalam tinjauan pustaka juga berisi mengenai hasil akhir/kesimpulan dari masing-masing jurnal/*paper* tersebut. Landasan teori berisi mengenai konsep, teori maupun rumus-rumus yang mendukung proses penelitian.

Bab 3, Perancangan sistem, bab ini membahas rancangan sistem yang dibangun mulai dari spesifikasi sistem, rancangan diagram sistem, rancangan antar-muka sistem dan tahap-tahapan yang berkaitan dengan proses dan pembuatan sistem tersebut.

Bab 4, Implementasi dan analisis sistem, bab ini akan menguraikan hasil implementasi dari metode-metode yang digunakan pada penelitian penulis dan analisis sistem secara teoritis berdasarkan *confusion matrix*.

Bab 5, Kesimpulan dan saran, bab ini akan membahas mengenai hasil analisis dari penelitian yang sudah dilakukan oleh penulis. Penulis juga akan memberikan saran yang mendukung supaya penelitian tersebut menjadi lebih baik. Dan diharapkan dari saran tersebut dapat memperbaiki kinerja sistem tersebut.

BAB 5
KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil analisis penelitian yang telah dibahas pada Bab 4, maka dapat ditarik kesimpulan sebagai berikut:

1. Berdasarkan analisa penulis, dapat disimpulkan penerapan *stopword filtering* pada sistem ekstraksi informasi sudah tepat.
2. Kekurangan dari *stopword filtering* adalah menghapus *bigram* yang mengandung kata-kata terdiri dari 1 kata nama makanan dan 1 kata *stopword*. 1 kata nama makanan tersebut merupakan nama makanan yang hanya tediri dari 1 kata.
3. Sistem ini cukup baik dalam mengekstraksi nama makanan dengan rata-rata nilai akurasi hasil evaluasi sistem untuk 2 jenis percobaan yaitu dengan *stopword filtering* dan tanpa *stopword filtering* adalah 97%. Hal ini menunjukan bahwa metode *n-gram, lexicon model* dan aturan baik dalam mengenali nama-nama makanan nusantara.
4. Indonesia memiliki sekitar 5.350 resep masakan (Indonesian Embassy, 2011), sedangkan penulis mengumpulkan sekitar 1500 nama makanan. Oleh karena itu sistem butuh lebih banyak nama makanan dan artikel makanan agar dapat mengekstraksi lebih banyak nama makanan.
5. Sistem belum bisa menyelesaikan dengan cepat dokumen yang berisi lebih dari 1000 kata dikarenakan pemrosesan yang lama. Oleh karena itu penulis meyimpulkan bahwa perlu adanya penyederhanaan *coding* dan spesifikasi perangkat keras yang lebih tinggi dari yang telah ditetapkan penulis untuk memproses dokumen uji lebih cepat lagi.

6. Sistem tidak dapat memproses kata dalam nama makanan yang memiliki makna lain selain nama makanan. Hal ini membuat sistem mengekstraksi nama makanan yang tidak sesuai dengan konteks dalam kalimat.

5.2. Saran

Sistem ini sangat memungkinkan untuk dilakukan pengembangan lebih lanjut sesuai kebutuhan yang terus bertambah, sehingga dapat meningkatkan akurasi sistem. Saran yang diajukan penulis dalam pengembangan sistem kedepannya adalah sebagai berikut:

1. Indonesia memiliki sekitar 5.350 resep masakan (Indonesian Embassy, 2011), sedangkan penulis mengumpulkan sekitar 1500 nama makanan. Penulis menyarankan untuk melengkapi daftar nama-nama makanan yang terdapat di dalam leksikon yang digunakan. Sehingga sebuah kata dapat diprediksi adalah bagian dari nama makanan.
2. Berdasarkan poin 1 penulis menyarankan untuk melengkapi data latih yang membahas nama makanan yang belum didapatkan oleh penulis. Sehingga dapat memberikan nilai frekuensi yang tepat untuk kata yang memang merupakan nama makanan.
3. Untuk mengatasi kekurangan *stopword filtering*, sistem perlu ditambahkan pemrosesan filtering *bigram* yang mengandung 1 kata nama makanan dan 1 kata *stopword*. 1 kata nama makanan tersebut adalah nama makanan yang memang hanya terdiri dari 1 kata. Proses tambahan ini berguna untuk menyimpan nama makanan tersebut sehingga tidak dibuang oleh proses *stopword filtering*.
4. Pada penelitian ini, Ekstraksi Informasi dilakukan hanya untuk mengekstraksi nama makanan nusantara. Penulis mengharapkan agar program terus dikembangkan, seperti ditambahkan fitur analisa sentimen konsumen terhadap makanan, apakah positif atau negatif dan fitur *web crawling* yang dapat mencari artikel yang membahas makanan dalam *website* dan *blog* untuk mendapatkan informasi nama makanan yang sedang banyak dibahas saat itu.

DAFTAR PUSTAKA

24/7 Wall St. (2010, September 30). *The Top Ten Ways Workers Waste Time Online*. Retrieved September 15, 2016, from http://247wallst.com

Agarwal, R., Miller, K. (2011). *Information Extraction from Recipes*. Stanford University.

Albukhitan, S., Helmy, T. (2013). *Automatic Ontology-Based Annotation of Food, Nutrition and Health Arabic Web Content*. King Fahd University of Petroleum & Minerals.

Bovée, C. L., & Thill, J. V. (2016). *Business communication today*. Upper Saddle River, NJ: Prentice Hall.

Chopra, D., Joshi, N., & Mathur, I. (2016). *Mastering Natural Language Processing with Python*. Packt Publishing Ltd. ISBN 178398905X, 9781783989058

Craig, T. (2008). *How to Avoid Information Overload*. Personnel Today.

Day, B. (2015, May 13). *About Indonesian food*. Retrieved September 15, 2016, from http://www.sbs.com.au/food/article/2008/07/01/about-indonesian food

Hanna, B. (2009). *Business Social Media Benchmarking Study*. Business.com.

Haryawan, C. (2014, February 2). *Pemanfaatan Sparql Inferencing Notation (Spin) Dalam Prototipe Pencarian Data Restoran Berbasis Semantik*. Jurnal Teknologi Technoscientia, *6*, 2nd ser., 110-122. ISSN:1979-8415

Herdiawan (2015). *Analisis Sentimen Terhadap TELKOM Indihome berdasarkan Opini Publik menggunakan Metode Improveed K-Nearest Neighbor.*

Indonesian Embassy (2011). *Wonderful Taste of Indonesia*. Retrieved July 29, 2018, from http://indonesianembassy.org.uk/wonderful-taste-of-indonesia

Indriati, Ridok, A. (2016). *Sentiment Analysis for Review Mobile Applications Using Neighbor Method Weighted K-Nearest Neighbor (NWKNN).* Journal of Environmental Engineering & Sustainable Technology. Malang.

Jurafsky, D., & Martin, J. (2000). *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics.* New Jersey: Alan Apt.

Machado, T. *Information Extraction – Automatic Recipe Introduction According to an Ontology*. INESC ID Lisboa.

Novianti, K.D.P, Setiawan, N.A., Kusumawardani, S.S. (2015). *Peningkatan nilai recall dan precision pada Penelusuran Informasi Pustaka Berbasis Semantik*. Jurnal Proceedings Konferensi Nasional Sistem dan Informatika (KNS&I).Universitas Gajah Mada.

Pusat Bahasa (2008). *Kamus Besar Bahasa Indonesia Edisi Keempat*. Jakarta: Gramedia Pustaka Utama. ISBN 9789792238419.

Scheffer, T., Wrobel, S., Popov, B., Ognianiv, D., Decomain, C., & Hoche, S. (2002, May 25). Learning Hidden Markov Models for Information Extration A tively from Partially Labeled Text. *Künstliche Intelligenz, 2*, 1-9. doi:10.1.1.1.9739

Sugianto, S. A. et all, (2013). *Pembuatan Aplikasi Predictive Text Menggunakan Metode N-Gram-Based*. Universitas Kristen Petra.

Suhartono, D., Christiandy, D.,Rolando (2013). *Lemmatization Technique In Bahasa: Indonesian Language.* Kuwait: Journal of Software (unpublished)

Susanti, E. & Mustofa, K. (2015). *Ekstraksi Informasi Halaman Web Menggunakan Pendekatan Bootstrapping pada Ontology-Based Information Extraction*. Universitas Gajah Mada.

Tala, F. Z. (2003). *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. M.S. thesis. M.Sc. Thesis. Master of Logic Project. Institute for Logic, Language and Computation. Universiteti van Amsterdam The Netherlands.

Triawati, Chandra (2009). *Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia.* Institut Teknologi Bandung.

Ueta, T., Iwakami, M., Ito, T. *A Recipe Recommendation System based on Automatic Nutrition Information Extraction.* Nagoya Institute of Technology.